

Validity, Responsiveness, and Minimal Important Difference for the SF-6D Health Utility Scale in a Spinal Cord Injured Population

Bonsan Bonne Lee, MBBS, FAFRM, MMED (Clin Epi), MHA, Gcert (IT),^{1,2,3}

Madeleine T. King, BSc, GradDipMedStats, PhD,⁴ Judy M. Simpson, PhD, CStat,²

Mark J. Haran, MBBS, FAFRM, MMed (Clin Epi), MHA,⁵ Martin R. Stockler, MBBS MSc FRACP,²

Obaydullah Marial, MPH, Gdip (IT), Gcer (IT),¹ Glenn Salkeld, B.Bus, G Dip Health Economics, MPH, PhD²

¹Prince of Wales Spinal Unit, Prince of Wales Hospital, Sydney, NSW, Australia; ²University of Sydney, Sydney, NSW, Australia; ³University of New South Wales, Sydney, NSW, Australia; ⁴Centre for Health Economics Research and Evaluation, University of Technology, Sydney, NSW, Australia; ⁵Royal North Shore Hospital, Sydney, NSW, Australia

ABSTRACT

Objective: To determine the feasibility, acceptability, discriminative validity, responsiveness, and minimal important difference (MID) of the SF-6D for people with spinal cord injury (SCI).

Methods: A total of 305 people with SCI completed the SF-36 health status questionnaire at baseline and at subsequent occurrence of a urinary tract infection (UTI) or 6-month follow-up. Normative SF-36 data were obtained from the Australian Bureau of Statistics. SF-36 scores were transformed to SF-6D utility values using Brazier's algorithm. We used UTI as the external criterion of clinically important change to determine responsiveness and two categories of the SF-36 transition question ("somewhat worse" and "somewhat better") as the external criterion to determine the MID. Derived SF-12 responsiveness was also assessed.

Results: The mean SF-6D values were: 0.68 (SD 0.21, n = 305) all patients; 0.66 (SD 0.19, n = 167) tetraplegia; 0.72 (SD 0.26, n = 138) paraplegia; 0.57 (SD 0.15, n = 138)

with UTI. The Australian normative SF-6D mean value was 0.80 (SD 0.14, n = 18,005). The SF-6D was able to discriminate between SCI and the Australian normative sample (effect size [ES] = 0.86), tetraplegia–paraplegia (ES = 0.23), and it was responsive to UTI (ES = 0.86 SF-36 variant, ES = 0.92 SF-12 variant). The MID for respondents who reported being somewhat worse or somewhat better at follow-up was 0.03 (SD 0.17, n = 108/305), while the MID for only those who were somewhat worse was 0.10 (SD 0.14, n = 58).

Conclusions: The content of the SF-6D is more appropriate than that of the SF-36 for this physically impaired population. The SF-6D has discriminative power and is responsive to clinically important change because of UTI. The MID is consistent with published estimates for other disease groups.

Keywords: health economics, health preference state, SF-6D, spinal cord injury, urinary tract infection, validity.

Introduction

Preference-based measures of health allow the relative value of health states to be compared, both within and across diseases [1]. A fundamental concept underpinning this is health utility [2], a measure of preference for health outcomes. Combined with survival data, utilities can be used to estimate quality-adjusted life-years (QALYs). Utilities and QALYs are used in cost-utility analyses to assess the relative value of health interventions, across a range of purposes (preventive, diagnostic, curative, palliative), types (programs, services, technologies, pharmaceuticals), and populations (within and across diseased, disabled, and healthy

populations). Preference-based measures are therefore useful and important outcome measures for policy-makers, both locally and internationally.

The SF-6D, a relatively new utility measure, is particularly attractive as it is calculated from the SF-36, a health status measure commonly used to assess the impact of disease and disability, including spinal injury [3]. In common with other multiattribute utility instruments such as EQ-5D (Euroqol) and Health Utilities Index (HUI) [4], it allows those experiencing the health states to contribute directly to utility scores. A particular advantage of the SF-36 and SF-6D is that they economize on data collection, yielding measures of both health status and utility. Since the SF-6D methodology was published 5 years ago, it has rapidly become a popular method of utility estimation. A recent systematic review of the use of health status measurement instruments to calculate QALYs found

Address correspondence to: Bonsan Bonne Lee, PO Box 431, Broadway, Sydney, NSW 2007, Australia. E-mail: blee@medicalinformatics.net
10.1111/j.1524-4733.2007.00311.x

that, despite its contemporary origin, the SF-6D accounted for 5% of the instruments used [5].

The SF-6D is a utility measure based on a six-dimensional health state classification. It is derived from a subset of 11 SF-36 questions covering the dimensions of Physical Functioning, Role Limitation, Social Functioning, Pain, Mental Health, and Vitality. It allows a possible 18,000 health states to be defined. A survey (involving SF-6D, ranking, and standard gamble) of 249 health states defined by the SF-6D was valued by a representative sample of the UK general public ($n = 611$). Econometric methods were then used to determine a model for predicting the standard gamble scores generated by the valuation survey [6]. Brazier et al. have shown that the SF-6D is a viable alternative preference measure [6]. It can be derived from either the SF-36 [6] or the shorter SF-12 [7]. It has been suggested that the SF-6D may be more sensitive than the EQ-5D, especially for mild-moderate health issues [6]. Limitations and outstanding issues with the SF-6D include whether it compromises the richness of the original SF-36 [6] and whether it is less sensitive when used in poorer health states [6,8,9]. It is therefore important that additional validation studies are performed in different populations and settings. The present article describes such a validation in an Australian population with spinal cord injury (SCI), most of whom were living in a general community setting.

The minimal important difference (MID) allows clinicians to determine whether a change observed on a self-reported health rating scale is meaningful or trivial. It has been defined as the smallest difference in score that the patient perceives as beneficial [10]. For our purposes, in the absence of significant side effects or cost barriers, this would lead to a change in clinical decision-making.

This article provides the first validation and MID values for the SF-6D in the SCI population. We assess the acceptability and appropriateness of the SF-6D for application in SCI, evaluate its discriminative ability, and determine its responsiveness to clinically important change. The external criterion used to define clinically important change is the occurrence of a urinary tract infection (UTI), a common comorbidity in this patient population group, with a reported incidence of 1.82 episodes per annum [11,12].

Methods

Data were collected during the Spinal-Injured Neuropathic Bladder Antisepsis randomized controlled trial [13]. Subjects were sampled from the New South Wales (State) Spinal Cord Injuries Database [14] and related databases of two major teaching hospitals. Inclusion criteria were: SCI with neurogenic bladder; stable bladder management; absence of serious renal

pathology; not taking antibiotics at enrollment; and absence of symptoms of a UTI at enrollment. Between November 2000 and August 2002, 543 eligible patients (predominantly community dwelling) were invited to participate in the study, of whom 305 (56%) agreed. Characteristics of the sample and reasons for nonparticipation are reported elsewhere [13].

Subjects completed the SF-36 at enrollment and again on development of UTI. If no UTI was experienced, a repeat SF-36 was completed at 6-month follow-up. Subjects completed the SF-36 by self-report with a research officer present, or by self-report via mail. Incomplete responses or inconsistencies were clarified by direct inquiry. Interpreters and physical assistance were used where necessary.

The SF-6D utility and dimensional scores were derived from SF-36 responses using Brazier's algorithm [6]. The domains and SF-36 items [15] used to construct the SF-6D were: Physical Functioning (items 3a, 3b, and 3j); Role Limitation due to physical problems (item 4c) and Role Limitation due to emotional problems (item 5b); Social Functioning (item 10); both bodily pain items (items 7 and 8); Mental Health (items 9b and 9f); and Vitality (item 9e). To explore whether the SF-12 version of the SF-6D differed in responsiveness from the full SF-36 version, SF-6D utility scores were recalculated using the Brazier SF-12 algorithm [7] for the responsiveness analysis.

Acceptability and appropriateness were assessed in terms of feasibility and content validity. Practical difficulties in the use of the SF-36 or content issues identified by subjects, research assistants, or authors during administration were recorded. Ceiling and floor effects for each SF-6D domain were examined by neurological level of injury (tetraplegia vs. paraplegia) [16].

Discriminative validity was assessed with cross-sectional comparisons of mean SF-6D utility and dimensional scores, externally by comparing the SCI patients with Australian normative data, and internally by comparing various subgroups of the SCI patients. Normative Australian SF-36 data were from 18,005 respondents in the Australian Bureau of Statistics National Health Survey of 1995 [17]. Means and standard deviations for normative and sample data were adjusted to fit the age and sex distribution of the Australian population using direct standardization [18].

Internal cross-sectional comparisons were based on a priori hypotheses made by three experts, one in rehabilitation medicine (BL) and two in quality-of-life research (MK, MS), who independently ranked their expectations about the size and direction of differences in SF-36 scales between groups defined by six clinically relevant variables. When applied to the single SF-6D index score, these led to six a priori hypotheses about the derived utility scores: that more (vs. less) extensive neurological level, more (vs. less) completeness of

injury, older (vs. younger) age, unemployment (vs. employment), being female (vs. male), and more (vs. less) recent injury would be associated with lower (vs. higher) utility scores. Further, the experts predicted that the mean differences in SF-6D utility scores between groups dichotomized by neurological level, completeness of injury, age, and employment status would be larger than those between groups dichotomized by sex and time since injury, and that effect sizes (ES) for the former set of characteristics would be small to moderate (0.2–0.5) while those for the latter would be at best small (<0.2).

Mean differences in SF-6D scores between groups were tested using *t*-tests for independent samples. For between-group comparisons, ES were calculated following the method of Kazis et al. [19]: $ES = (m_1 - m_2) / s_1$, where m_1 is reference group mean, m_2 is comparison mean, and s_1 is reference group standard deviation.

The responsiveness of the SF-6D to clinically relevant change in health status was determined in the 138 patients who developed a UTI during the course of the clinical trial. The mean change in the SF-6D utility score from the first to the second assessment was calculated for scores based on both the SF-36 and SF-12 algorithms. For within-group comparisons, we used the longitudinal form of ES (mean change/standard deviation of change), also called the standardized response mean [7].

The MID was calculated using the method of Walters and Brazier as follows [20]. The health transition item (Question 2) of the SF-36 (which is not part of the SF-6D) was used to define respondents who had experienced a MID. This question asks “Compared to 1 year ago, how would you rate your health in general

now?”, with five response options: 1 = much worse, 2 = somewhat worse, 3 = about the same, 4 = somewhat better, and 5 = much better. A score of “2” or “4” was deemed equivalent to the MID. Where patients reported a worsening of health, the sign of the SF-6D score change was reversed before combining with those patients who reported improvement. Our methodology differed from Walter and Brazier’s in one crucial respect: our questionnaire used the original SF-36 Question 2, which compares health now to 1 year ago, whereas they modified the comparison time frame to be consistent with the duration of follow-up [20].

The trial was approved by the ethics committees of the participating hospitals (Royal North Shore Hospital, Prince of Wales Hospital, Prince Henry Hospital, and Royal Rehabilitation Center Sydney). All computations and statistical tests were conducted in SAS version 9 (SAS Institute Inc., Cary, NC, USA).

Results

Trial participants had a longer mean time since SCI (by 2.6 years) than those who were excluded or did not consent. There were no other significant differences between participants and nonparticipants [13].

Participants had a mean age of 44 years (SD 14, range 16–82 years) and were predominantly male (83%). Fifty-five percent of patients had tetraplegia and 49% had a complete spinal injury. The median time since SCI was 12 years (range from 1 month to 61 years). The characteristics used to define subgroups for assessing discriminative validity are presented in Table 1.

Table 1 Baseline characteristics of 305 patients with spinal cord injury categorized by criteria used to assess discriminative validity, mean SF-6D scores for each subgroup, differences between groups, and effect sizes

| Characteristics | n | % | Mean* (SD) | Difference (P-value [†]) | Effect size [‡] |
|------------------------------------|-----|----|-------------|------------------------------------|--------------------------|
| Neurological level | | | | | |
| Tetraplegia | 167 | 55 | 0.68 (0.13) | 0.05 (0.002) | 0.34 |
| Paraplegia [§] | 138 | 45 | 0.73 (0.15) | | |
| Completeness | | | | | |
| Complete (ASIA A) | 148 | 49 | 0.72 (0.15) | −0.02 (0.18) | −0.16 |
| Incomplete (ASIA B–D) [§] | 157 | 51 | 0.69 (0.14) | | |
| Age (year) | | | | | |
| 16–43 [§] | 157 | 51 | 0.70 (0.15) | −0.01 (0.83) | −0.02 |
| 44+ | 148 | 49 | 0.71 (0.14) | | |
| Employment (hour) | | | | | |
| No paid | 197 | 65 | 0.68 (0.15) | 0.05 (0.002) | 0.40 |
| Any paid [§] | 108 | 35 | 0.74 (0.14) | | |
| Sex | | | | | |
| Male [§] | 252 | 83 | 0.71 (0.14) | 0.04 (0.08) | 0.27 |
| Female | 53 | 17 | 0.67 (0.15) | | |
| Time since injury (year) | | | | | |
| ≤4 years | 90 | 30 | 0.68 (0.15) | 0.04 (0.04) | 0.26 |
| >4 years [§] | 215 | 70 | 0.71 (0.14) | | |

*Mean SF-6D score for each subgroup at recruitment, not adjusted to the age and sex distribution of the Australian population. Higher scores reflect better health.

[†]*t*-test for independent samples.

[‡]Effect size calculated following the method of Kazis et al. [19].

[§]Reference Group.

ASIA, American Spinal Injuries Association Neurological Classification [16].

Feasibility and Content Validity

There were no missing data. Issues about the content of the SF-36 raised by participants included: uncertainty if limitations of activities referred to a comparison with a non-SCI person or the patient's usual activities; problems with physical activity questions (particularly those involving walking or climbing stairs), and uncertainty about the period of recall required.

Ceiling and Floor Effects

The full range of levels was observed for all six dimensions of the SF-6D. Table 2a shows the results for the overall sample and stratified by neurological level. An apparent floor effect (37%) in the Physical Functioning dimension in the overall sample was accounted for almost entirely by the tetraplegia group (63%). In contrast, there were no notable floor effects in the Australian sample (Table 2b). Ceiling effects in the SCI sample (27–56%) occurred in the Role Limitation, Social Functioning, Pain, and Mental Health dimensions, irrespective of neurological category. Ceiling effects were also apparent in the normative sample, exceeding those in the SCI sample for Role Limitation (70% vs. 55%) and Social Functioning (63% vs.

50%), but interestingly less so for Mental Health (20% vs. 35%).

Discriminative Validity

Figure 1 shows mean scores for each of the six SF-6D dimensions for subjects with SCI (tetraplegia and paraplegia), compared to a large normative Australian population sample. The ES for overall utility (Table 3) express the degree of discrimination in the SF-6D index scores between the Australian norms and the SCI sample ($ES = 0.86$, $P < 0.0001$) and between paraplegic and tetraplegic subgroups ($ES = 0.23$, $P = 0.025$). Dissecting the SF-6D index into its six component domains (Table 3), we see the most pronounced differences in the Physical Functioning dimension, with paraplegic patients being clearly worse than the normative sample and the tetraplegic patients being demonstrably the worst group. The other dimensions (Role Limitation, Social Functioning, Pain, Mental Health, and Vitality) showed relatively small absolute differences and ES compared to the Physical Functioning dimension.

Table 1 shows the mean SF-6D utility scores for groups dichotomized by six clinically relevant baseline characteristics. Differences in mean SF-6D scores were in the direction expected with the exception of com-

Table 2 Proportion of responses (%) on each level* of the SF-6D dimensions in (a) the sample of SCI individuals and (b) the Australian normative sample (n = 18,005)

| SF-6D dimension levels | Physical Functioning | Role Limitation | Social Functioning | Pain | Mental Health | Vitality |
|--|----------------------|-----------------|--------------------|------|---------------|----------|
| <i>(a) Sample of SCI individuals</i> | | | | | | |
| All SCI (n = 305) | | | | | | |
| 1 | 6 | 55 | 50 | 36 | 35 | 5 |
| 2 | 9 | 25 | 18 | 20 | 26 | 40 |
| 3 | 14 | 7 | 12 | 18 | 27 | 31 |
| 4 | 24 | 13 | 12 | 13 | 10 | 14 |
| 5 | 10 | — | 8 | 10 | 2 | 10 |
| 6 | 37 | — | — | 4 | — | — |
| Tetraplegia (n = 167) | | | | | | |
| 1 | 2 | 54 | 51 | 43 | 34 | 4 |
| 2 | 1 | 27 | 18 | 19 | 27 | 40 |
| 3 | 5 | 8 | 13 | 14 | 28 | 28 |
| 4 | 20 | 11 | 11 | 11 | 9 | 16 |
| 5 | 10 | — | 7 | 11 | 2 | 12 |
| 6 | 63 | — | — | 2 | — | — |
| Paraplegia (n = 138) | | | | | | |
| 1 | 11 | 56 | 50 | 27 | 38 | 6 |
| 2 | 18 | 22 | 18 | 21 | 24 | 41 |
| 3 | 26 | 6 | 10 | 22 | 26 | 36 |
| 4 | 30 | 17 | 13 | 16 | 11 | 12 |
| 5 | 10 | — | 9 | 9 | 1 | 7 |
| 6 | 5 | — | — | 6 | — | — |
| <i>(b) Australian normative sample</i> | | | | | | |
| 1 | 45 | 70 | 63 | 39 | 20 | 8 |
| 2 | 30 | 10 | 19 | 25 | 38 | 54 |
| 3 | 13 | 11 | 12 | 21 | 31 | 27 |
| 4 | 6 | 9 | 5 | 8 | 8 | 8 |
| 5 | 3 | — | 2 | 6 | 2 | 3 |
| 6 | 3 | — | — | 2 | — | — |

*Two dimensions have six levels (Physical Functioning, Pain); three dimensions have five levels (Social Functioning, Mental Health, Vitality); one dimension has only four levels (Role Limitation). Level 1 represents the best health (ceiling) and the highest level represents the worst health (floor).

SCI, spinal cord injury.

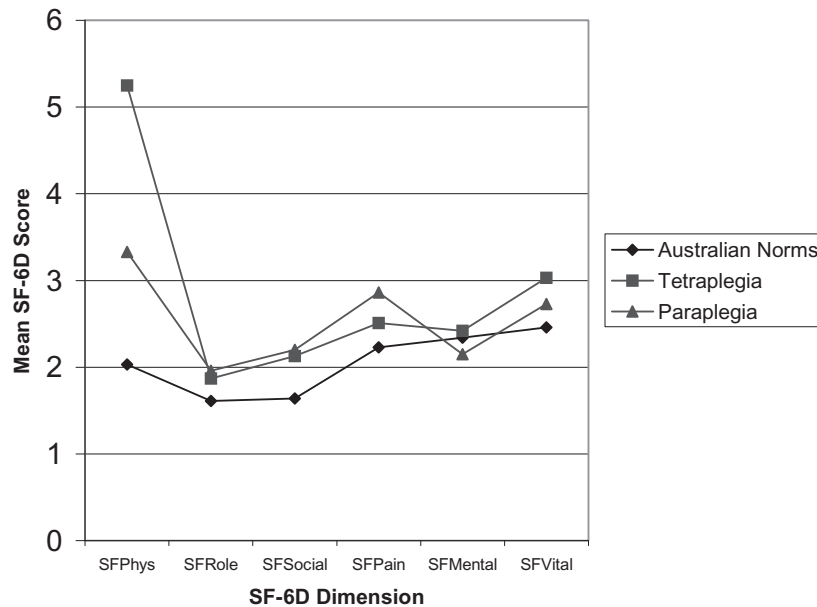


Figure 1 SF-6D profiles for tetraplegic ($n = 138$) and paraplegic patients ($n = 167$) compared to a normative Australian population sample ($n = 18,005$): all values are means adjusted for the age and sex distribution of the Australian population. A higher score represents worse health. The range of scores is as follows: SFPhys (Physical Functioning) 1–6, SFRole (Role Limitation) 1–4, SFSocial (Social Functioning) 1–5, SFPain (Pain) 1–6, SFMental (Mental Health) 1–5, SFVital (Vitality) 1–5.

pleteness of injury and age. ES were generally as expected, although those for sex and time since injury were larger than expected and those for age and completeness of injury were smaller than expected. Statistically significant differences were found only for neurological level, employment, and time since injury.

Responsiveness

For the 138 trial participants who developed UTI, the overall age- and sex-adjusted utility score was 0.68 (SD 0.20) before and 0.57 (SD 0.15) after UTI. Table 4 presents unadjusted results for those who developed

UTI, which demonstrate the responsiveness of the SF-6D to this common and clinically significant comorbidity. Corresponding results for the 167 trial participants who did not develop UTI show the extent to which SF-6D utility scores change over a 6-month period in a SCI sample whose health is stable (Table 4). The larger absolute differences and ES in the former group and the small absolute differences and ES in the latter group confirm the responsiveness of SF-6D. Figure 2 decomposes the utility index into its six component domains, showing that the subjects who developed UTI reported significantly worse SF-6D scores in

Table 3 SF-6D scores for (a) a normative Australian population sample ($n = 18,005$) and the SCI sample ($n = 305$) and (b) a paraplegic sample ($n = 138$) and a tetraplegic sample ($n = 167$): overall utility values* and individual dimension scores† adjusted for age and sex‡

| (a) SF-6D | Australian normative sample [§] Mean (SD) | SCI—all neurological levels Mean (SD) | Effect size |
|------------------|---|--|---------------------------|
| Overall utility* | 0.80 (0.14) | 0.68 (0.21) | 0.86 |
| SFPhysical | 2.03 (1.20) | 4.34 (1.94) | 1.93 |
| SFRole | 1.61 (1.01) | 1.91 (1.44) | 0.30 |
| SFSocial | 1.64 (0.99) | 2.17 (2.03) | 0.54 |
| SFPain | 2.23 (1.28) | 2.70 (2.24) | 0.20 |
| SFMental | 2.34 (0.97) | 2.32 (1.76) | -0.02 |
| SFVitality | 2.46 (0.87) | 2.90 (1.57) | 0.51 |
| (b) SF-6D | SCI—paraplegic sample [§] Mean (SD) | SCI—tetraplegic sample Mean (SD) | Effect size |
| Overall utility* | 0.72 (0.26) | 0.66 (0.19) | 0.23 |
| SFPhysical | 3.33 (1.72) | 5.25 (1.54) | 1.12 |
| SFRole | 1.96 (1.80) | 1.87 (1.44) | -0.05 |
| SFSocial | 2.20 (2.23) | 2.13 (2.09) | -0.03 |
| SFPain | 2.86 (2.66) | 2.51 (2.17) | -0.13 |
| SFMental | 2.15 (1.76) | 2.42 (1.61) | 0.15 |
| SFVitality | 2.73 (1.65) | 3.03 (1.67) | 0.18 |

*Overall utility: higher score is better (where 1 = perfect health and 0 = death). Significance test for overall utility: 2 sample t-test $P < 0.0001$ (a) and $P = 0.025$ (b).

†SF-6D mean dimension levels: higher score is worse.

‡Adjusted to the age and sex distribution of the Australian population.

§Reference group.

||Effect size calculated following the method of Kazis et al. [19].

Table 4 Responsiveness of SF-6D* assessed in subjects before and after developing a UTI; subjects who did not develop UTI during the study serve as controls

| | Developed UTI (n = 138) | | Did not develop UTI (n = 167) | |
|--------------------------|-------------------------|-------------------------|-------------------------------|-------------------------|
| | SF-12 [†] (SD) | SF-36 [‡] (SD) | SF-12 [†] (SD) | SF-36 [‡] (SD) |
| Mean SF-6D score* | | | | |
| Baseline [§] | 0.72 (0.14) | 0.70 (0.14) | 0.72 (0.14) | 0.71 (0.15) |
| Follow-up | 0.60 (0.12) | 0.58 (0.12) | 0.70 (0.14) | 0.68 (0.15) |
| Mean change in SF-6D | -0.12 (0.13) | -0.12 (0.14) | -0.02 (0.14) | -0.03 (0.14) |
| Effect size [¶] | 0.92 | 0.86 | 0.14 | 0.21 |

*Overall utility: higher score is better (1 = perfect health, 0 = death); not adjusted to the age and sex distribution of the Australian population.

[†]SF-6D derived from SF-12 scoring algorithm.

[‡]SF-6D derived from SF-36 scoring algorithm.

[§]Baseline assessments were at recruitment all subjects.

^{||}For subjects who developed UTI, follow-up assessments were completed on development of UTI. For subjects who did not develop UTI, follow-up assessments were completed 6 months after recruitment.

[¶]Longitudinal effect size for change is the standardized response mean.

UTI, urinary tract infection.

every dimension except physical after UTI relative to before they developed a UTI. Table 4 also shows that SF-6D utility values calculated using the Brazier SF-12 algorithm are as responsive as those calculated using the Brazier SF-36 algorithm.

Minimal Important Difference

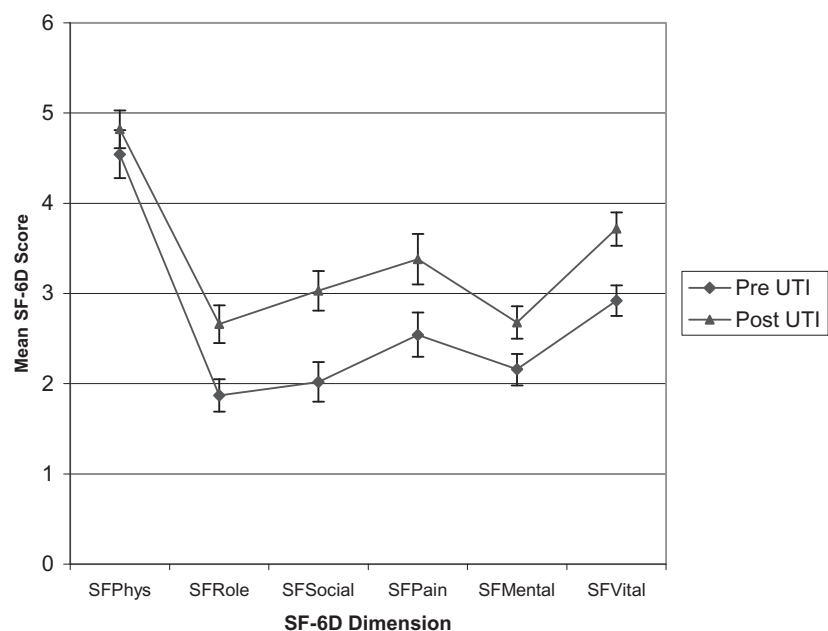
The MID had a mean value of 0.03 (n = 108; SD 0.17). When limited to those who were somewhat better, the MID was -0.04 (n = 50; SD 0.16), and when limited to those who were somewhat worse, the MID was 0.10 (n = 58; SD 0.14). The difference between the MIDs of these two groups was statistically significant ($P < 0.0001$). Of the group who were somewhat worse, most of the patients (39/58) suffered a UTI. In interpreting these MIDs, we note that patients who developed a UTI had a mean follow-up period of 64 days, compared to 182 days for those who did not

develop a UTI; these are less than the 1-year recall period specified in the SF-36 health transition question used as the external criterion for determining the MID.

Discussion

This report validates the SF-6D in a group of patients with SCI, predominantly living in the general community and participating in a randomized trial. It documents utility values and comparisons with a normative population, and demonstrates that both the SF-36 and SF-12 variants of the SF-6D were responsive to clinically important changes in disease state in a group with severe physical impairment. The SF-36 data were easily collected by a research assistant. The exclusion of several SF-36 physical activity questions, particularly those about walking or climbing stairs, which are problematic for people with SCI, may make the SF-6D

Figure 2 Responsiveness: SF-6D profiles for spinal cord injured patients (n = 138) before and after developing urinary tract infection (UTI): all results are mean values for each dimension with 95% confidence intervals displayed, adjusted for the age and sex distribution of the Australian population. A higher score represents worse health. The range of scores is as follows: SFPhys (Physical Functioning) 1–6, SFRole (Role Limitation) 1–4, SFSocial (Social Functioning) 1–5, SFPain (Pain) 1–6, SFMental (Mental Health) 1–5, SFVital (Vitality) 1–5.



more acceptable to the SCI population. The SF-6D was able to discriminate between the SCI sample and the general population and also between tetraplegic and paraplegic patients, but it was unable to detect the differences among people with SCI according to neurological completeness.

The strengths of this study include the perfect completion rate of the study instrument and the well-defined cohort sampled from a comprehensive register. A limitation of this study is that the sample was derived from a randomized controlled trial targeting participants with SCI and neurogenic bladder. Compared to the overall New South Wales SCI population, sample characteristics which may bias toward poorer utility assessments include the trial's inclusion criterion of neurogenic bladder [21] and older age. On the other hand, the community-based nature of this sample may bias the results toward better utility scores. Employment levels (35%) were within the published range for this population group [22]. In common with other clinical trials, voluntary participants tend to be more motivated than nonparticipants [23].

The SF-6D differs markedly from the underlying SF-36, in that the physical function questions that relate specifically to walking and stair climbing are omitted (SF-36 items 6, 7, 9–11). This redresses a major problem of the SF-36 in this patient population, where these questions are seen to be ambiguous or irrelevant for SCI individuals. SCI researchers are therefore justified in questioning the appropriateness and validity of the standard SF-36 for assessing the health status in this population [3,24–26]. Although the SF-6D seems to overcome this issue, the problem remains whether to obtain data for the SF-6D via the standard SF-36. There are several responses to this problem: researchers interested in collecting the SF-6D could use the standard SF-36 or a modified version of it (such as modifications of Meyers, Andresen, or Tate [26–28]), do both (as we did) or just apply the subset of questions used to derive the SF-6D dimension levels and final utility score. Meyers and Tate suggest replacing the words “walk” and “climb” with the word “go,” arguing that this wording allows SCI individuals to take into account assistive equipment, while maintaining adequate construct validity [27]. Provided that the modification chosen does not affect the calculation of the SF-6D, researchers may find a modified SF-36 more acceptable. Using both standard and modified scales (in essence asking the problematic physical questions twice, with modified wording to maintain compatibility) may make people less sensitive to the problematic questions as they at least have a relevant response to one of two related questions.

As people adapt to disability, they reconceptualize their reference of comparison for health states, resulting in higher self-reported health ratings than expected [29]. Utility scores taken soon after the traumatic event

that causes a SCI are likely to be very different to those taken many years later, as in the majority of our study sample. When comparisons were made between the SCI and normative samples, the better than expected mental dimension (and the little changed role, social, and vitality scores) for the SF-6D probably reflects this response shift (Fig. 1). Within the SCI sample (Table 1), time since spinal injury was significantly associated with SF-6D utility scores, with a moderate ES that was larger than expected. As expected, subjects with longer time since injury had higher utilities than those more recently injured, which in part may reflect postaccident adjustment. Such response shift effects are likely to lead to the mean utility score improving as time since SCI increases, which could attenuate the effect of any disease state deterioration.

The Australian normative values calculated to compare discriminative validity revealed an overall age- and sex-adjusted utility score of 0.80 (SD 0.14) from a sample of 18,005 participants. This is very similar to the value derived by Bharmal and Thomas [30] in a sample of 11,248 North American participants: 0.81 (SD 0.18).

When SF-12 and SF-36 variants of the SF-6D were compared, they yielded similar utility values. This suggests that the SF-36 and SF-12 variants of the SF-6D are equally responsive, and that the SF-6D of either derivation is suitable for detecting the clinical change resulting from a UTI in a population with SCI.

In dissecting the SF-6D index into its six constituent domains, the Physical Functioning dimension was the major discriminator between tetraplegia, paraplegia, and the Australian normative sample, with relatively little difference seen for the other SF-6D dimensions (Fig. 1). In contrast, the Physical Functioning dimension was the least responsive to UTI disease state change, with relatively large differences being detected in the other domains: Role Limitation, Social Functioning, Pain, Mental Health, and Vitality (Fig. 2). This is expected as UTI would not be likely to shift significantly a physical score already subject to floor effects, particularly in the 55% of participants with tetraplegia. In contrast, the other dimensions either had ceiling effects or had room to deteriorate in response to a UTI. Given that the SF-6D is designed to be used as a single index of utility, this differential response across the dimensions is not a problem, and may indeed allow flexibility of use in other populations.

Walters and Brazier found that the SF-6D MID for 11 disease groups ranged from 0.011 to 0.097 with a mean of 0.041. The MID for SCI with UTI disease state change from this study (0.03, SD 0.17) is consistent with these previous estimates. It is important to note that the results from the two groups used to calculate the MID (“somewhat worse” and “somewhat better”) were significantly different in our sample. Our questionnaire used the original SF-36

Question 2, which compares health now with 1 year ago, while Walters and Brazier modified the time frame in this question to be consistent with the duration of follow-up [20]. One of the 11 samples reported by Walters and Brazier likewise showed a significant difference between the “somewhat worse” and “somewhat better” groups. This highlights a potential problem in using Question 2 from the SF-36 for calculating the MID, as combining results for the two groups assumes they are identical except for sign [20]. Several issues may contribute to our result: the appropriateness of the retrospective SF-36 Question 2 as a valid external criterion of change in health state, the mismatch of time periods in the retrospective and prospective SF-36 assessment in this study, and the effect of response shift. Nevertheless, this article provides the first published estimates of MID for the SCI population using Walter and Brazier’s methodology.

There is debate as to whether preference weights should be sourced from community samples or directly from participants in a clinical trial [5,31]. Problems can arise when community preference weights attribute artificially low utility scores to disease groups like SCI, which may discriminate against people with SCI by valuing their years of life less highly, and therefore valuing health interventions for them less favorably. On the other hand, response shift in our directly measured sample probably contributed to higher than expected values in some dimensions. It is important that additional studies with longitudinal data are performed within the SCI population to further clarify responsiveness and the effects of response shift, and to compare SCI utility values derived from SF-6D and other methods with those of other population groups.

Conclusion

The SF-6D can reliably discriminate not only the gross differences between persons with SCI and a normative population group, but also the smaller and more clinically relevant differences between patients with paraplegia and tetraplegia. The SF-12 and SF-36 variants of the SF-6D are both responsive to the additional disease burden of UTI in this patient group. The MID is consistent with published estimates for other disease groups. The content of the SF-6D makes it more appropriate than the SF-36 for use in this physically impaired population.

Sources of financial support: Motor Accidents Authority, New South Wales (Financial), Brucia Pharmaceutical (nonfinancial).

SF-36 Health Survey © 1988, 2002 by Medical Outcomes Trust and QualityMetric Incorporated. All rights reserved. SF-36 is a registered trademark of the Medical Outcomes Trust.

References

- 1 Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health* 2000;21:587–611.
- 2 Drummond MF, O’Brien BSG, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press, 1997.
- 3 Wood-Dauphinee S, Exner G, Bostanci B, et al. Quality of life in patients with spinal cord injury—basic issues, assessment, and recommendations. *Restor Neurol Neurosci* 2002;20:135–49.
- 4 Kopeck JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol* 2003;56:317–25.
- 5 Rasanen P, Roine E, Sintonen H, et al. Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review. *Int J Technol Assess Health Care* 2006;22:235–41.
- 6 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271–92.
- 7 Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;42:851–9.
- 8 Bryan S, Longworth L. Measuring health-related utility: why the disparity between EQ-5D and SF-6D? *Euro J Health Econ* 2005;6:253–60.
- 9 Longworth L, Bryan S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Econ* 2003;12:1061–7.
- 10 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. *Control Clin Trials* 1989;10:407–15.
- 11 Waites KB, Canupp KC, Devivo MJ. Epidemiology and risk factors for urinary tract infection following spinal cord injury. *Arch Phys Med Rehabil* 1993;74:691–5.
- 12 Cardenas DD, Hoffman JM, Kirshblum S, McKinley W. Etiology and incidence of rehospitalization after traumatic spinal cord injury: a multicenter analysis. *Arch Phys Med Rehabil* 2004;85:1757–63.
- 13 Lee BB, Haran MJ, Hunt LM, et al. Spinal-injured neuropathic bladder antiseptics (SINBA) Trial. *Spinal Cord* 2007;45:542–50.
- 14 Lee BB, Middleton JW, Rutkowski SB, Tagney B. The introduction of the NSW spinal injuries database system. *Proceedings of Paralympics Scientific Conference, Sydney, Australia, 2000*.
- 15 Ware JE Jr, Gandek B. Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. *J Clin Epidemiol* 1998;51:903–12.
- 16 American Spinal Injury Association. *International Standards for Neurological Classification of Spinal Cord Injury*. Chicago: American Spinal Injury Association, 2000.
- 17 Australian Bureau of Statistics. *National Health Survey, Australia 1995; Confidentialised Unit Record File: Technical Paper*. Canberra: Australian Bureau of Statistics, 2001.
- 18 Armitage P, Berry G. *Standardisation*. In: Armitage, PBG, ed. *Statistical Methods in Medical Research*. Oxford: Blackwell Science, 1994.

- 19 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27(Suppl.):S178–89.
- 20 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523–32.
- 21 Westgren N, Levi R. Quality of life and traumatic spinal injury. *Arch Phys Med Rehabil* 1998;79:1433–9.
- 22 Kennedy P, Lude P, Taylor N. Quality of life, social participation, appraisals and coping post spinal cord injury: a review of four community samples. *Spinal Cord* 2006;44:95–105.
- 23 Holden G, Rosenberg G, Barker K, et al. The recruitment of research participants: a review. *Soc Work Health Care* 1993;19:1–44.
- 24 Andresen EM, Fouts BS, Romeis JC, Brownson CA. Performance of health-related quality-of-life instruments in a spinal cord injured population. *Arch Phys Med Rehabil* 1999;80:877–84.
- 25 Dijkers M. Quality of life of individuals with spinal cord injury: a review of conceptualisation, measurement, and research findings. *J Rehabil Res Dev* 2005;42:87–110.
- 26 Meyers AR, Andresen EM. Enabling our instruments: accommodation, universal design, and access to participation in research. *Arch Phys Med Rehabil* 2000;81(Suppl.):S5–9.
- 27 Tate DG, Kalpakjian CZ, Forchheimer MB. Quality of life issues in individuals with spinal cord injury. *Arch Phys Med Rehabil* 2002;83(Suppl.):S18–25.
- 28 Andresen EM, Meyers AR. health-related quality of life outcomes measures. *Arch Phys Med Rehabil* 2000;81(Suppl.):S30–45.
- 29 Groot W. Adaptation and scale reference bias in self-assessments of quality of life. *J Health Econ* 2000; 71:403–20.
- 30 Bharmal M, Thomas J 3rd. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the us general population. *Value Health* 2006;9:262–71.
- 31 Hays RD, Hahn H, Marshall G. Use of the SF-36 and other health-related quality of life measures to assess persons with disabilities. *Arch Phys Med Rehabil* 2002;83(Suppl.):S4–9.